

# DESPERATELY SEEKING NON-GAUSSIANITY

*The Light Curve of 0957+561*

W.H. PRESS AND G.B. RYBICKI

*Harvard-Smithsonian Center for Astrophysics  
Cambridge, MA 02138 (USA)*

## Abstract.

“Non-Gaussian” is the casual explanation often given for anything unexpected in an astronomical time series. What better place to look for non-Gaussianity, therefore, than in the light curve of 0957+561, the gravitational lens that, until recently, had yielded frustratingly inconsistent determinations of its lag. We discuss the difficulties in measuring deviations from Gaussianity in weakly nonstationary processes (such as  $1/f$  noise or random walk) and define a restricted set of “well-behaved” three-point statistics. An important special case of such a well-behaved statistic is the skew of a linear combination of the data, with coefficients summing to zero. Analytic and Monte Carlo calculations evaluate the performance of such a statistic in the case of a non-Gaussian “wedge model” (shot noise, with each shot having a rapid rise and slow decay). We find that even for as well studied an object as 0957+561, the detectability of any deviation from Gaussian is problematical at best. At present, one can rule out a wedge model only if the individual shots are as infrequent as one in 10-20 days.

## 1. Introduction

By now it is well established, most recently and definitively by Kundić et al. (1996, hereafter “K96”), that the time delay of the lensed quasar 0957+561 is around 420 days, confirming previous determinations by Vanderriest et al. (1989); Pelt et al. (1996), who used data by R. Schild and D. Thomson; and others. Not surprisingly, it is a matter of concern to the present authors that the longer values, around 540 days, obtained by Press, Rybicki, and Hewitt (1992a,b, hereafter “PRH”), using a method based on unbiased Wiener filtering (details in Rybicki and Press 1992, hereafter “RP”) has proved to be wrong.

We still know of nothing wrong with the method described in PRH and RP. Indeed, when applied to the new data of K96, the method readily finds the (correct) 420 day delay; in this we confirm the independent analysis reported in K96. Furthermore, the method is known mathematically to be in some sense optimal for data generated by a Gaussian process. Extensive, and successful, Monte Carlo simulations were also reported in PRH. So, it is something of a mystery why PRH failed, rather flamboyantly, on the 0957+561 data sets that were originally tried.

In situations like this, the epithet “non-Gaussian” is frequently heard. Indeed it is well known that the application of methods validated (by theorem or Monte Carlo) on Gaussian processes to non-Gaussian ones can sometimes lead to wrong answers – though, more often, to correct answers but underestimated error bars. The present situation thus gives us a good excuse for thinking about how to measure or characterize non-Gaussianity, and an occasion to search for non-Gaussianity in the 0957+561 light curve.

## 2. The Kundić et al. Data Set

K96 reports the results of two seasons’ observations of 0957+561 A,B. Since lens component B is delayed, one effectively gets a free third season of light curve for the underlying quasar. We have used the method described in RP (1) to bring the A and B components to a common flux scale (undoing the lens magnification ratio), and (2) to construct an optimal reconstruction of the light curve, along with error bars on the optimal reconstruction. Figure 1 shows the K96 data points, with measurement uncertainties, and the reconstruction and its  $1\text{-}\sigma$  error range. Note that the error properly balloons out in the interseasonal gaps, and is properly smaller than the individual measurement errors when the density of measurements is large enough to allow, in effect, the averaging of nearby points.

The sharp drop that occurs around JD 2449700, first reported by Kundić et al. (1995), is the feature that makes the 420 day time delay as unambiguous as one might like. One might be tempted to guess that a feature like this is evidence of non-Gaussianity, i.e., is unlikely in a purely Gaussian process with the time spectrum (or correlation function) of 0957+561. But how is one to know *quantitatively* whether this is the case?

The simplest possible question, and also the question that involves the lowest-order moment (or cumulant) that can deviate from Gaussianity, is to ask: Is there any statistically sound evidence that the light curve in Figure 1 is time-asymmetric? For example, can we substantiate that the data favors rapid declines (and more gradual increases) over rapid increases (and more gradual declines)? This question can in principle be answered by a *three-point statistic*.

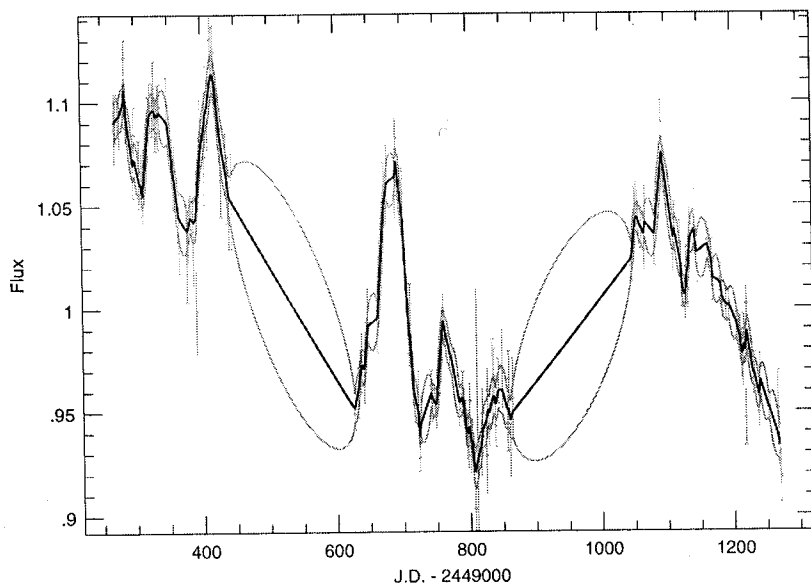


Figure 1. Light curve of 0957+561 in arbitrary flux units (g band). The data points and errorbars derive from Kundić et al. (1996). The heavy curve is the unbiased Wiener optimal reconstruction. The light curves are the  $1\text{-}\sigma$  error range of the reconstruction.

If we fail to find time asymmetry, the next obvious question will be: Are the declines and increases *both* sharper (or less sharp) than predicted by a Gaussian process. This question can only be answered by a statistic that probes *four-point* behavior. As we shall see, it is correspondingly harder to answer with limited data.

### 3. Weakly Nonstationary Processes

We face the handicap that the 0957+561 light curve, like that of most quasars, is a “weakly nonstationary” process, in the sense that its power spectrum diverges at low frequencies  $f$  faster than  $1/f$ . This, we will see below, renders most standard three-point statistics inapplicable or unreliable, and gives our problem a different cast from previous searches for time-asymmetry in astronomical time series (e.g., Weisskopf et al. 1978).

Suppose we measure some process  $h(t)$  for only a total time  $T$ , and then estimate its mean  $\bar{h}$  by some average of the measured values. Using Parseval’s theorem and the convolution theorem, it is easy to show that the

variance of  $h$  around its apparent mean is

$$\langle (h - \bar{h})^2 \rangle \approx \int_{2\pi/T}^{\infty} P(f) df \quad (1)$$

where  $P(f)$  is the power spectrum. If the integral diverges as  $T$  is increased, then the variance diverges. Operationally, one would find no tendency for the estimates  $\bar{h}$  to converge as  $T$  increases. We should regard the mean of a process with divergent variance as “unknowable or infinite”.

If we parametrize the low-frequency power spectrum by a single exponent  $\beta$ ,

$$P(f) \sim f^\beta \quad (2)$$

then a process is weakly nonstationary for  $1 \leq \beta < 3$ . Quasars and AGNs usually have  $\beta$  in the range of about 1.5 to 2. The case  $\beta = 1$  is called “1/f noise”, while  $\beta = 2$  characterizes a random-walk process.

One easily relates this parametrization of the power spectrum to an equivalent model for the autocorrelation function or structure function. If the correlation decays as a power law in lag,

$$\langle h(t)h(t + \tau) \rangle = B - \frac{1}{2}b\tau^{\beta-1} \quad (3)$$

for some constants  $B$ ,  $b$ , and  $\beta$ , and in some range  $0 \leq \tau < \tau_{max}$ , so that the structure function is

$$\langle [h(t + \tau) - h(t)]^2 \rangle = b\tau^{\beta-1} \quad (4)$$

then the Wiener-Khinchine theorem gives (in an appropriate range of frequencies)

$$P(f) \approx \frac{b\Gamma(\beta) \sin[\frac{1}{2}(\beta - 1)\pi]}{(2\pi f)^\beta} \quad (5)$$

for  $1 < \beta < 3$ , showing that the  $\beta$ 's in equations (2), (3) and (4) are indeed the same parameter. Notice, for later comment, that a structure function that increases linearly with  $\tau$ ,  $\beta = 2$ , corresponds to a random walk power spectrum.

#### 4. “Well-Behaved” Three Point Statistics

Given a set of measured values  $y_i \equiv y(t_i)$ ,  $i = 1, \dots, N$ , the most general three-point statistic that we can write is

$$S = \sum_{ijk} \alpha_{ijk} y_i y_j y_k \quad (6)$$

for some particular choice of the three-index kernel  $\alpha_{ijk}$ . Indeed, one can regard *all* three-point higher order statistics, or estimators of such statistics over a data set, as being simply particular choices for  $\alpha_{ijk}$ . Without loss of generality, one can take  $\alpha_{ijk}$  to be symmetrical on all its indices, since in equation (6) it is contracted with a symmetrical combination of  $y$ 's. All statistics like this vanish (in expectation value) for a Gaussian process of zero mean, so a statistically significant nonzero value is always evidence of non-Gaussianity. However, different ones of these statistics can have vastly different *variances* when applied to a purely Gaussian signal. It is this statistical variance that limits the sensitivity of any particular statistic in distinguishing a Gaussian from a non-Gaussian process, especially, as we now show, for weakly nonstationary processes.

Starting with equation (6), we can write

$$\text{Var}(S) = \langle S^2 \rangle = \sum_{ijklmn} \alpha_{ijk} \alpha_{lmn} \langle y_i y_j y_k y_l y_m y_n \rangle \quad (7)$$

For a Gaussian process, the expectation value of the sixth order product is equal to the sum of 15 terms, each the product of three second order products and each with coefficient unity. If the  $y_i$ 's have the autocorrelation given by the model equation (3), then it is not hard to show that the sixth order expectation has the form

$$\begin{aligned} \langle y_i y_j y_k y_l y_m y_n \rangle = & 15B^3 - 3B^2[b_{ij} + b_{ik} + \text{more terms}] \\ & + B[b_{kl}b_{mn} + \text{many more terms}] + O(B^3) \end{aligned} \quad (8)$$

Recall now that for a weakly nonstationary process the value  $B$  (total variance) is infinite. So, we'd better try to choose  $\alpha_{ijk}$ 's such that the terms of order  $B^3$ ,  $B^2$ , and  $B$  exactly cancel. Remarkably it is possible to do just this. The condition on the  $\alpha_{ijk}$ 's that results is

$$\sum_i \alpha_{ijk} = \sum_j \alpha_{ijk} = \sum_k \alpha_{ijk} = 0 \quad (9)$$

An important special case is where the  $\alpha_{ijk}$ 's are the sum of a number of "rank one" symmetric pieces,

$$\alpha_{ijk} = \sum_q a_q \gamma_{qi} \gamma_{qj} \gamma_{qk}, \quad \text{with} \quad 0 = \sum_i \gamma_{qi} \quad \text{for all } q \quad (10)$$

Here there can be any number of terms indexed by  $q$ , and the resulting statistic  $S$  is

$$S = \sum_q a_q \left( \sum_i \gamma_{qi} y_i \right)^3 \quad (11)$$

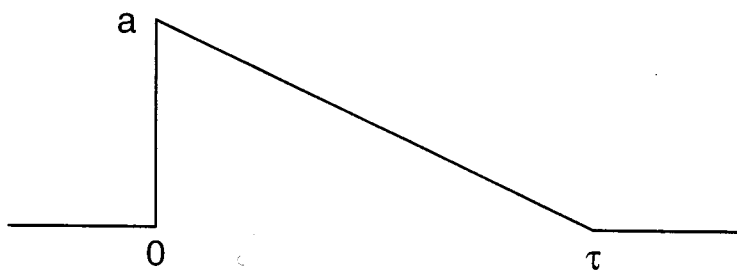


Figure 2. Basic shape for a shot-noise model with wedge-shaped "shots". The process is the sum of shots occurring at Poisson-random times with some rate  $\nu$ . This process is time-asymmetric and has a nonzero expectation for three-point statistics.

which is something like an estimator of the skew of a set of linear combinations of the data, each combination with coefficients summing to zero. (Although we present this as a special case of equation (9), we conjecture that this is actually the most general case, if there are a sufficient number of terms indexed by  $q$ .)

A further specialized case of this statistic is amenable to some analytical calculation in an instructive model case, which we will next do.

## 5. The Wedge Model

Although we can compute the variance of a statistic like equation (11) in the presence of Gaussian noise, we can't compute its expectation value in the presence of a non-Gaussian process – unless we have a specific model for that process. Let us take as such a model a shot-noise process with wedge-shaped shots of infinitely rapid rise and slow (extending over a time  $\tau$ ) decay. For simplicity all shots have the same amplitude  $a$ . They occur Poisson randomly at a mean rate  $\nu$ .

For any shot-noise process with shot shapes  $\psi(s)$ , the autocorrelation function and structure function can be shown to be

$$C(t) \equiv \langle (h(s) - \bar{h})(h(s+t) - \bar{h}) \rangle = \nu \int ds \psi(s+t)\psi(s) \quad (12)$$

$$S_2(t) \equiv \langle [h(s+t) - h(s)]^2 \rangle = 2[C(0) - C(t)] \quad (13)$$

while the skew of a lagged difference (a statistic of the form of equation (11), not coincidentally) can be shown to be

$$S_3(t) \equiv \langle [h(s+t) - h(s)]^3 \rangle = 3\nu \int ds \psi(t+s)\psi(s)[\psi(s) - \psi(t+s)] \quad (14)$$

For the wedge-shaped shots of Figure 2, one can do the integrals, yielding

$$C(t) = a^2 \nu \tau \frac{1}{3} \left(1 - \frac{t}{\tau}\right)^2 \left(1 + \frac{t}{2\tau}\right) \quad (|t| < \tau) \quad (15)$$

$$S_3(t) = a^3 \nu t \left(1 - \frac{t}{\tau}\right)^2 \left(1 + \frac{t}{2\tau}\right) \quad (|t| < \tau) \quad (16)$$

with  $C(t) = S_3(t) = 0$  for  $|t| \geq \tau$ . It is also instructive to give the results to lowest order in  $t/\tau$  for  $t \ll \tau$ , and also to include the effect of a Gaussian measurement error  $\sigma$  (assumed constant) for each measured value of  $y(t)$ :

$$S_2(t) = a^2 \nu t + 2\sigma^2 \equiv V't + 2\sigma^2, \quad S_3(t) = a^3 \nu t \quad (17)$$

(Being Gaussian,  $\sigma$  contributes nothing to  $S_3$ .) Notice the linear increase of the structure function  $S_2(t)$  with time. This shows that, for times  $t \ll \tau$ , the wedge model has the two-point statistics of a random walk (close to what is actually observed in 0957+561). We define  $V' \equiv a^2 \nu$  since only this combination, and not  $a$  or  $\nu$  separately, is observable in any two-point statistic.

The nondimensional skew, formed from  $S_2(t)$  and  $S_3(t)$  is

$$\text{Skew} = \frac{S_3(t)}{S_2(t)^{3/2}} = \frac{a^3 \nu t}{(a^2 \nu t + 2\sigma^2)^{3/2}} = \frac{1}{(\nu t)^{1/2} [1 + 2\sigma^2/(V't)]^{3/2}} \quad (18)$$

Notice that the skew goes linearly to zero at small lags  $t$ , because it is lost in the measurement error  $\sigma$ , and also goes to zero and large lags, inversely with  $\sqrt{\nu t}$ , the mean number of shots that occur in one lag time, an example of the central limit theorem in action.

A more elaborate calculation is required, however, to determine the *detectability* of the the skew: we need to know not only its expected value, but also its variance. For a specific calculation, we need also to define how the averaging in equation (14) is to be done. To this end, let  $\Delta(s, t) \equiv y(s+t) - y(s)$ , and

$$\widehat{S_3}(t) \equiv \frac{1}{T} \int_0^T \Delta^3(s, t) ds \quad (19)$$

where “hat” denotes an estimator and  $T$  is the total length of observation. This is an idealized model, because it replaces the finite set of data points  $y_i$  with a continuous observable  $y(t)$ ; we will comment further on this below. We now calculate

$$\text{Var}(\widehat{S_3}(t)) = \frac{1}{T^2} \int_0^T \int_0^T du dv \langle \Delta^3(u, t) \Delta^3(v, t) \rangle \quad (20)$$

Making the approximation that we can factor the sixth order moment *as if* it were Gaussian (essentially the central limit theorem), a lengthy calculation gives

$$\text{Var}(\widehat{S_3(t)}) = 3\sqrt{2}\frac{t}{T}(V't)^3 \left[ 3 \left( 1 + \frac{2\sigma^2}{V't} \right)^2 + 1 \right] \quad (21)$$

Now combining equations (17) and (21), we get the “detectability of the skew” (or of  $S_3$ ) measured in standard deviations,

$$\frac{\text{Skew}}{\sqrt{\text{Var}(\text{Skew})}} \approx \frac{1}{3 \cdot 2^{1/4}} \frac{1}{(\nu t)^{1/2}} \left( \frac{T}{t} \right)^{1/2} \frac{1}{1 + 2\sigma^2/(V't)} \quad (22)$$

Interestingly, this does not go to zero for small  $t$ , but is rather monotonically decreasing with increasing  $t$ , with maximum value

$$0.140(T\nu)^{1/2} \frac{V'}{\nu\sigma^2} \quad (23)$$

at  $t = 0$ . The reason that  $t = 0$  is preferred here, but not in equation (18), is our assumption, above, of continuously measurable data, so that  $y(t)$  can be estimated much more accurately than the  $\sigma$  of a single measurement. We will see now that this assumption is actually not too bad for the K96 data set!

## 6. Application to 0957+561

From the K96 data, all of the quantities in expression (23) are known except  $\nu$ :  $T \sim 400$  days (total length of observation, exclusive of the interseasonal gaps),  $\sigma \sim 0.01$  (that is, measurement accuracies of about 0.01 mag),  $V' = 2 \times 10^{-5} \text{ day}^{-1}$  (estimating the structure function from the data). Setting expression (23) to 2 for a  $2\text{-}\sigma$  detection, we find that the wedge model's skew should be evident in the data only if the mean rate of shots is less than about  $0.08 \text{ day}^{-1}$ , i.e., less than one shot every 12 days – even though the data set's observations are typically only a day or two apart. This is sobering, and shows the inherent difficulty of detecting non-Gaussianity in noisy data, even several seasons of astronomically high quality.

To see whether these analytic estimates, which involved a number of idealizations, are correct, we have performed Monte Carlo simulations, as follows:

First, we generate many synthetic realizations of the 0957+561 data, each with the same times of observation and measurement errors as the actual K96 data set, and each drawn from a distribution with the same

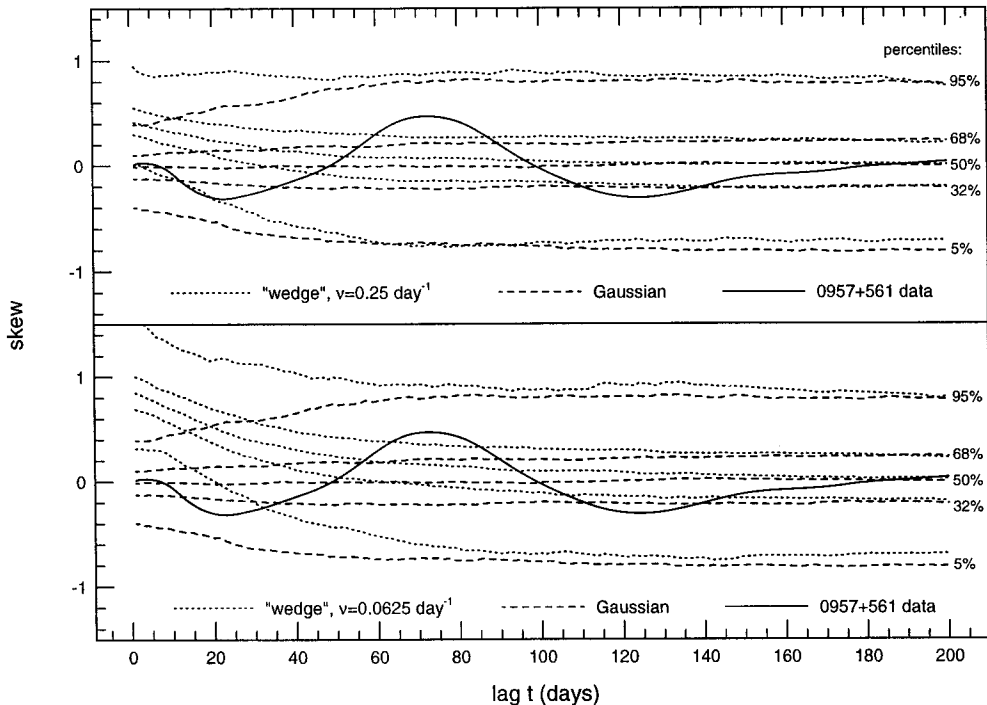
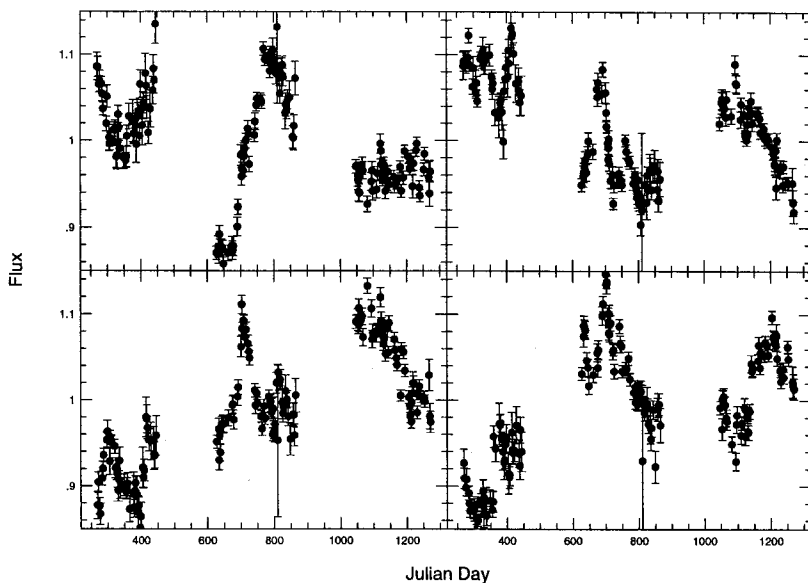


Figure 3. Results of Monte Carlo experiments comparing the skew of “wedge” and Gaussian models with the same 2-point properties as the 0957+561 actual data. Dotted and dashed curves are percentile results of many realizations. At small lags the wedge model with a rate  $\nu = 0.0625 \text{ day}^{-1}$  (lower panel) is easily distinguished from Gaussian, while the wedge model with a rate  $\nu = 0.25 \text{ day}^{-1}$  (upper panel) is not. The solid curve (both panels) is derived from the K96 data for 0957+561, and is clearly not distinguishable from the Gaussian model. The lower panel wedge model is ruled out by this data.

two-point statistics (correlation function or power spectrum) as the K96 data. Some realizations are generated as a purely Gaussian process, while others embody the “wedge model” already described. In the latter case, we always take  $\tau = 200$  days for the decay time (our results are highly insensitive to this parameter), but we try several different values of  $\nu$  (the mean rate).

Second, for each realization (and for the actual data also) we apply the machinery described in PRH and RP to get an unbiased Wiener filter reconstruction of the underlying light curve as a continuous function (conceptually at least), and its error bars.

Third, we estimate the skew of the difference of this continuous function at various lags  $t$ , essentially the statistic  $\widehat{S}_3(t)/S_2(t)^{3/2}$  of equation (19). (A slight modification is that we use the reconstruction error bars to de-weight



*Figure 4.* Three Monte-Carlo realizations, along with the actual 0957+561 data. The realizations all have times of observation and measurement errors identical to the actual data, and have the same two-point statistics as the data. One realization is purely Gaussian. The other two are “wedge models” with respective rates of  $0.25 \text{ day}^{-1}$  and  $0.0625 \text{ day}^{-1}$ . The statistical test described in the text easily identifies the latter process as non-Gaussian. Can the reader tell which is which? [Answer in text.]

strongly lags with ends that fall in the interseasonal gaps in the data.)

Fourth, after accumulating many synthetic realizations for each choice of parameters, we plot the percentile confidence intervals of the skew statistic as a function of lag  $t$ . For the actual data we simply plot, on the same scale, its skew statistic.

Results of this procedure for the Gaussian case and wedge models with two choices of  $\nu$ , are shown in Figure 3. One sees that at large lags, the wedge models are indistinguishable from Gaussian (identical percentile intervals). At small lags, their non-Gaussianity is to some extent detectable. For  $\nu = 0.25 \text{ day}^{-1}$  the detectability is not certain, since there is substantial overlap of plausible confidence intervals. For  $\nu = 0.0625 \text{ day}^{-1}$ , however, the detectability is quite reliable since (e.g.) the 95% percentile curve of the Gaussian model reaches only to the 5% curve of the wedge model. In general, the Monte Carlo results substantiate the analytic estimate of equation (22). This shows that the K96 data is dense enough in time that the analytic assumption of a continuous function was justified.

Thus for detectability; now, what about actual detection in the real data? One sees that, for all lags, the skew of the actual data is definitively within the range of a Gaussian model, and at or outside the 5% confidence bound for the triangle models for all lags less than 20 days ( $\nu = 0.25 \text{ day}^{-1}$ ) or 30 days ( $\nu = 0.0625 \text{ day}^{-1}$ ). These models are thus excluded at the 5% level. Indeed at small lags, where a skew signal should be strongest, the data is (coincidentally) at the 50% percentile of the Gaussian model. We should note that while the wedge model makes zero lag ( $t = 0$ ) the most sensitive indicator, a model with a finite rise time would suppress the skew statistic for times smaller than that rise time, so the exclusion of wedge models at finite lag also serves to exclude models with finite rise times.

Notice that the time-reversed models, with slow rise and rapid quenching (for which the dotted curves in Figure 3 would be flipped in sign), are less strongly excluded. We doubt that this is in any way significant, however.

## 7. Conclusions

Non-Gaussianity, even when quite extreme (as in the wedge model) can be quite hard to detect in data of quantity and quality comparable to K96 – which, by astronomical standards, is very good data indeed! Of course, one can get lucky: a single  $20\text{-}\sigma$  flare in the data would be definitively non-Gaussian. Teasing non-Gaussianity out of a signal that is starting to satisfy the central limit theorem (shot noise, e.g.) is what is difficult. If you still don't think so, look at Figure 4, which shows typical realizations of the processes already described. [Answers: the upper-left panel is Gaussian; upper-right is actual data; lower-left is a wedge model with  $\nu = 0.0625 \text{ day}^{-1}$  (easily detectable by the statistic discussed in this paper); lower-right has  $\nu = 0.25 \text{ day}^{-1}$  (not so easily detectable).]

The detectability of such non-Gaussianity increases only slowly with total observing time  $T$ , as the square root; but it can increase rapidly with decreasing measurement errors  $\sigma$  (cf. equations 18 and 22). However there is an minimum  $\sigma$ , with  $\sigma^2 \sim V't$ , after which further improvement is small. For 0957+561, this minimum  $\sigma$  is about  $0.003\sqrt{t_{\text{days}}}$  magnitudes, for lag times  $t$ .

Although we have treated only the 3-point skew here, it is clear from rough analytical calculations that analogs of equations (18) and (22) also hold for higher order statistics, e.g. kurtosis; these have less-favorable coefficients, and also higher powers of  $1 + 2\sigma^2/(V't)$  in the denominators. Skew (if it exists in the data) is in some sense the *most* detectable non-Gaussian statistic.

Although the limits are not very impressive, the K96 data, with the anal-

ysis of this paper, does rule out some shot-noise models for the 0957+561. At the 5% significance level, models with shots that have rise times of less than about 10–20 days, and mean shot rates of less than  $0.25 \text{ day}^{-1}$ , are excluded.

## References

- Kundić, T., Colley, W.N., Gott, J.R., Malhotra, S., Pen, U., Rhoads, J.E., Stanek, K.Z., and Turner, E.L. (1995) *Astrophys. J. (Lett.)*, **455**, L5.
- Kundić, T., Turner, E.L., Colley, W.N., Gott, J.R., Rhoads, J.E., Wang, Y., Bergeron, L.E., Gloria, K.A., Long, D.C., Malhotra, S., and Wambsganss, J. (1996) preprint astro-ph/9610162 [K96].
- Pelt, J., Kayser, R., Refsdal, S., and Schramm, T. (1996) *Astron. Astrophys.*, **305**, 97–106.
- Press, W.H., Rybicki, G.B., and Hewitt, J.N. (1992) *Astrophys. J.*, **385**, 404 [PRH<sub>1</sub>].
- Press, W.H., Rybicki, G.B., and Hewitt, J.N. (1992) *Astrophys. J.*, **385**, 416 [PRH<sub>2</sub>].
- Rybicki, G.B. and Press, W.H. (1992) *Astrophys. J.*, **398**, 169 [RP].
- Vanderriest, C., Schneider, J., Herpe, G., Chevreton, M., Moles, M., and Wlerick, G. (1989) *Astron. Astrophys.*, **215**, 1–13.
- Weisskopf, M.C., Sutherland, P.G., Katz, J.I., and Canizares, C.R. (1978) *Astrophys. J. (Lett.)*, **223**, L17–L20.